Recovery of 3D Pose of Bones in Single 2D X-ray Images

Piyush Kanti Bhunre Wee Kheng Leow
Dept. of Computer Science
National University of Singapore
3 Science Drive 2, Singapore 117543
{piyushka, leowwk}@comp.nus.edu.sg

Tet Sen Howe
Dept. of Orthopaedics
Singapore General Hospital
Outram Road, Singapore 169608
tshowe@sgh.com.sg

Abstract

Estimation of the 3D pose of bones in 2D images plays an important role in computer-assisted diagnosis and surgery. Existing work has focused on registering a 3D model of the bone to the images of the same bone. This approach incurs a high health care cost on the patients, and are prescribed only for certain procedures. Such computer assistance is not routinely available to diagnostic and surgical procedures that involve only x-ray images, limiting the precision of the procedures. This paper proposes a method for recovering 3D pose by registering a generic 3D model of a bone to single x-ray images of different patients. It can perform 3D-2D registration with unknown correspondence automatically. Comprehensive test results show that the method is accurate and robust in recovering the pose of femurs in the x-ray images of different patients.

1. Introduction

Estimation of the 3D pose of bones in 2D images plays an important role in computer-assisted diagnosis and surgery [2, 3, 4, 5, 6, 11]. Existing work has focused primarily on the technique of registering a 3D model of a bone to the images of the same bone, and recovering the 3D pose from the rigid transformation that yields the best registration. In the case of computer-assisted surgery, for example, the 3D model is often obtained from CT or MR scan of a patient, and it is registered to intra-operative x-ray images of the same patient.

This approach has a major limitation. CT and MR scans incur a high health care cost on the patients, and are prescribed only for certain procedures. Such computer assistance is not routinely available to diagnostic and surgical procedures that involve only x-ray images, limiting the precision of the procedures.

This paper describes a 3D pose estimation method that registers a generic 3D model of a bone to single x-ray images of different patients (Section 3). It does not require fiducial markers attached to the patients and it can perform 3D-2D registration with unknown correspondence automatically. Moreover, it is much faster than the methods that register digitally reconstructed radiographs (DRR) [4, 11] generated from CT or MR volume to the target images. Comprehensive tests were conducted to assess the performance of the method (Section 4). Quantitative test results show that the method is accurate and robust in recovering the pose of proximal and distal femurs in x-ray images of different patients. In contrast, most existing articles do not report quantitative test results.

2. Related Work

A Survey of medical image registration methods is given in [7]. Registration-based 3D pose estimation methods can be categorized into two classes: feature-based and appearance-based.

Feature-based methods [2, 3, 5] match geometric features extracted from images to those in the 3D model. The method of [2] represents a 3D model by its multiple 2D views. Each view is represented by a set of primitive geometric features of the bone's contour such as corner points and curve segments. 2D-2D matching is performed between the model views and a target image, and the best-matching view indicates the bone's pose. Thus, the method's accuracy is limited by the number of 2D views and the angular resolution of the views.

The methods of [3, 5] compute 3D lines back-projected from image contour points and minimize their signed distances from the 3D model's surface. Both methods represent a volumetric 3D model as an octree, which uses a significant amount of memory. The accuracy of the 3D representation depends on the spatial resolution of the octree. Computing signed distances of lines from model surface is a time-consuming operation. Guéziec et al. [3] improved the algorithms's efficiency by first segmenting the surface from the model.

Appearance-based methods [4, 9, 10, 11] synthesize a digitally reconstructed radiograph (DRR) from a patient's CT or MR volume. The matching error between the pixel intensities of the DRR and the patient's x-ray image is computed. By determining the orientation that yields the best match, the pose of the 3D model is recovered. Synthesizing DRR is a time-consuming process. LaRose [4] developed a more efficient method of generating DRR by directly programming the graphics card.

Appearance-based methods have the advantage of not requiring feature extraction from the target images. However, x-ray images often contain irrelevant contents (e.g., noise edges and surgical instruments) that can confuse these methods. Moreover, a good initialization is necessary for their successful application. Otherwise, they can be easily trapped in the local minima of the objective function during optimization. To mitigate the weakness, [6] synthesized gradient image by projecting the gradients of the CT volume onto the image plane, and performed feature-based registration before gradient-based registration.

The methods in [2, 3, 5] estimate femur pose whereas [4, 11] estimate pelvis pose. The methods in [2, 4] register 3D model to single image whereas [3, 5, 6, 11] register to two images. Only [3, 5, 9, 10, 11] have reported quantitative test results. All these methods used 3D model and 2D image of the same patient. In contrast, our method uses a single generic 3D model to estimate the 3D pose of femurs in the x-ray images of different patients. So, it has to be robust against shape variations across patients. To maximize accuracy and robustness, 3D femur is represented as point-and-mesh surface model, and is registered to the boundary contour of the femur in the x-ray image. Compared to volumetric model, point-and-mesh model requires less memory and can be acquired at very high resolution.

Note that the focus of our problem is to recover the pose (i.e., rotation angles) of the bone in 2D images of different patients. Hence, rigid registration is adopted. If non-rigid registration is used, the 3D model can be deformed sufficiently to yield a good match with the 2D image even if its pose is incorrect. All the methods discussed above perform rigid registration for the same reason.

3. 3D Pose Estimation by Registration

The inputs to the registration problem consist of the set M of 3D points on the surface mesh of the 3D model and the set C' of 2D image points on the image boundary contour. The set $C \subset M$ of 3D model points $\mathbf{X}$ that project to the bone contour in the image plane depends on the rotation $\mathbf{R}$, translation $\mathbf{T}$, and projection $\mathbf{P}$ of the 3D model. So, C is unknown and has to be determined by the registration algorithm, along with the correspondence $f : C \rightarrow C'$ between the two sets of points. Thus, 3D-2D registration with unknown correspondence can be formulated as the problem of determining the C, $\mathbf{R}$, $\mathbf{T}$, $\mathbf{P}$, and f that minimize the error E:

$$E = \frac{1}{|C|} \sum_{\mathbf{X} \in C} \|\mathbf{P}(\mathbf{R}\mathbf{X} + \mathbf{T}) - f(\mathbf{X})\|^2 . \qquad (1)$$

The optimal $\mathbf{R}$ gives the pose of the 3D model. Scaled-orthographic projection is used in our implementation because it simplifies the algorithm and the results obtained are accurate enough (Section 4). An iterative Newton's method of root finding is used to solve the problem. The algorithm can be summarized as follows:

3D Pose Estimation Algorithm

A. Extraction of image boundary contour (Section 3.1).

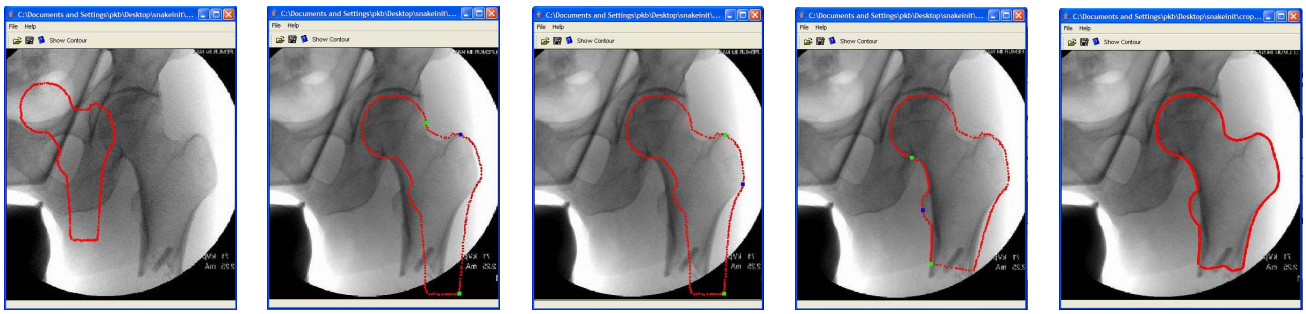
B. Automatic initialization:

1. Set initial fixed default pose (Fig. 3(D0) and Fig. 4(E0)).
2. Scale the 3D model so that the width of the bounding box of its 2D projection is equal to that of the image boundary contour.
3. Translate the 3D model so that the center of its 2D projection coincides with that of the image boundary contour.

C. Repeat until convergence:

1. Apply rigid transformation on the 3D model.
2. Project the 3D model onto the image plane.
3. Extract the boundary contour of the projected model silhouette.
4. Perform 2D-2D registration with unknown correspondence between model contour points and image contour points and obtain point correspondence between them (Section 3.2). This in turn establishes the correspondence between 3D model points and image contour points (Section 3.3).
5. Perform 3D-2D registration between 3D model points and image contour points (Section 3.3), and update rigid transformation parameters.

The algorithm is repeated until the error E does not change significantly. As will be shown in Section 4, the algorithm can converge to (local) optimal solutions for most test images. For a small number of cases in which the algorithm does not converge, it can be terminated after running for a fixed number of iterations. In all cases, the best registration result is kept at each iteration, and it is reported as the final registration solution when the algorithm terminates.



(a) (b) (c) (d) (e)

Figure 1. Image contour extraction with interactive initialization. Red curves denote the model contour, green dots denote fixed end points of contour segments, and blue dots denote movable feature points. (a) Initial random placement of model contour. (b–d) Placement of various parts of the model contour in sequence. (e) Final contour extracted by snake algorithm.

3.1. Image Contour Extraction

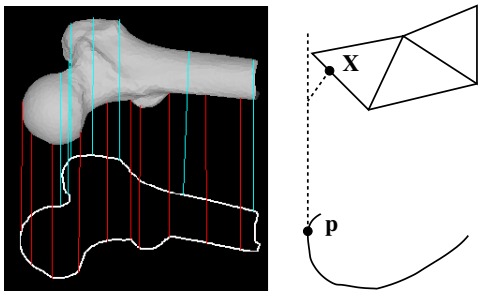
Accuracy of pose estimation depends on the accuracy of the extracted contour. A fully automatic boundary extraction method may have a wider applicability, but tends to be less accurate and robust. So, an easy-to-use semi-automatic method is developed to extract image boundary contour accurately.

A GUI is developed for the user to interactively initialize the segmentation algorithm. The user can easily drag, scale, rotate, and place a model contour to fit the contour in the image. Dragging (i.e., translation), scaling, and placement are achieved, respectively, by moving a computer mouse, turning the mouse wheel, and clicking the mouse button. As the various parts of the model contour are dragged and placed in sequence, they are automatically rotated (Fig. 1). The initialization process is done in five steps (Fig. 1), each step initializes a different part of the bone. The model contour is automatically deformed as different parts are localized in turn. After initialization, the snake algorithm is run to extract the image contour. A test performed on the femur images of 30 different patients yielded a mean error of 1.3 pixels and a standard deviation of 0.3 pixel compared to the ground truth contours extracted manually by experts. This result shows that the method is accurate in extracting femur contours in images.

3.2. 2D-2D Registration

2D-2D registration with unknown correspondence is solved using Iterative Closest Point (ICP) algorithm [1]. It seeks the 2D scale s', rotation matrix $\mathbf{R}'$, and translation $\mathbf{T}'$ that minimize the error E':

$$E' = \frac{1}{n}\sum_i \|s'\mathbf{R}'\mathbf{x}_i + \mathbf{T}' - f'(\mathbf{x}_i)\|^2 \qquad (2)$$



(a) (b)

Figure 2. Finding 3D-2D correspondence. A corresponding 3D model point is found by (a) back projecting a 2D model contour point p into 3D space, and (b) searching for the 3D point X on the edge of a triangular face that are nearest to the backprojection lines.

where $\mathbf{x}_i$ is a point on the projected model contour, f' is the closest-point function from the model contour points to the image contour points, and n is the number of points. Newton's method of root finding is applied to minimize E'.

3.3. 3D-2D Registration

The 2D-2D registration performed in the previous stage establishes a correspondence f' between the projected model contour points and image contour points. This in turn establishes the correspondence f between the 3D model points that project to the 2D model contour points and the image contour points. The 2D model contour is obtained by projecting and rendering 3D model using OpenGL library. It is necessary to determine the set C of 3D model points that project to the 2D model contour. Given a 2D model contour point, it is back-projected into 3D space and a 3D

point on an edge of a triangular face of the 3D model that is closest to the back-projection line is searched. It is a good approximation of the intersection point between the back-projected line (Fig. 2) and the 3D model surface. Search efficiency is improved by performing range search in 3D KD-tree [8]. Only a small number of distance computations is required to find the nearest point to the back-projection line.

In principle, when the correspondence f is known, it is possible to directly compute the transformation matrices $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{T}$. In our algorithm, the correspondence f' estimated by 2D-2D registration in the first iteration is only an initial guess. Just as for ICP algorithm [1], there are usually many wrong correspondences in the initial guess. So, instead of directly computing the transformation matrices, one optimization iteration is executed to update the transformation parameters. Then, the algorithm described in Section 3 is repeated. In this way, the optimal C, f, $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{T}$ are determined iteratively.

4. Performance Tests

A generic 3D model of the femur was constructed by scanning a real bone using a Minolta 3D scanner. Sixteen views of the bone were registered and merged using the mesh editing tool of Minolta scanner to form the complete 3D model. Comprehensive tests were performed to assess the algorithm's robustness and accuracy in recovering 3D pose from single input images.

4.1. Robustness Tests

Our pose estimation algorithm was applied to 198 proximal femur x-ray images of different patients obtained from a local hospital, with 172 (86.9%) healthy and 26 (13.1%) fractured femurs. Figure 3 illustrates sample registration results. The algorithm can register to images of healthy femur accurately, with matching error E ranging from 2 to 5, even when the pose after automatic initialization is quite different from the final registered pose (Fig. 3(H1, H2)). For images of fractured femurs, the algorithm can still register to them quite accurately as long as the femurs' shapes are not severely distorted (Fig. 3(F1, F2)) and the matching error ranges from 2 (for subtle fractures without shape change) to 9 (for moderate fractures with shape change). In the above cases, the matching error E always converges to stable values. For severely fractured femurs whose shapes are drastically distorted, the matching error tends to be larger and oscillates over iterations from as low as 7 to as high as 30. In summary, 87.4% of the test images have matching errors below 6. The mean, median, and standard deviation of errors are 3.85, 3.43, and 2.35 respectively.

Based on visual inspection of the registration results, we found that registration can be considered as successful for matching error E smaller than 9, which corresponds to an average error of 3 pixels between model contour points and image contour points. Based on this criterion, the registration success rate is 93.9%, which is much larger than the percentage of healthy femurs (86.9%) in the test images. This indicates that the algorithm is very robust against shape variations across patients and due to fractures.

The algorithm is also applied to fluoroscopic images of distal femurs (Fig. 4). The results show that the algorithm can also work well on a different type of bone whose shape is more regular than that of the proximal femur. Moreover, registration is successful even when part of the distal femur is occluded (Fig. 4(H5)).

4.2. Accuracy Tests

The methods for evaluating registration accuracy proposed in [9, 10] require that the 3D CT/MR volume and 2D x-ray images are taken from the same patient. In our case, patient-specific 3D models were not available because CT and MR were not prescribed to the patients. So, we generated 3D models based on real x-ray images of patients for evaluating our algorithm's accuracy. A similar method was adopted in [11] to measure the accuracy of 3D pose estimation. For pose estimation, we define the accuracy as the absolute difference between the 3D model's actual rotation angles and the estimated angles. Two types of tests were performed: single-patient test and cross-patient test.

Single-Patient Test: This test condition is the same as those in [2, 3, 4, 5, 6, 9, 10, 11]. That is, the 3D model used for pose estimation and the 2D images come from the same patient. The test was performed using generic 3D models of proximal and distal femurs. Given a 3D model, 100 synthetic 2D contours were generated by perspective projection of the models at a wide range of random orientations. Denote the angle of rotation about X-, Y-, and Z-axis as ϕ_x, ϕ_y, and ϕ_z. The image plane is defined as the X-Y plane, with Y-axis oriented along the femoral shaft, and Z-axis pointing out of the image plane. For proximal femur, the ranges of rotational angles were -6° to $+6^\circ$ for ϕ_x, 20° to 70° (Fig. 3(D1)) for ϕ_y, and -15° to $+15^\circ$ for ϕ_z, which were observed in medical practice. For distal femur, the ranges of the rotational angles were -6° to $+6^\circ$ for ϕ_x, -30° to 30° (Fig. 4(E1)) for ϕ_y, and -15° to $+15^\circ$ for ϕ_z as recommended by the doctor.

The algorithm successfully estimated the rotation angles of all the test cases. Table 1 shows the mean, median and standard deviation of the absolute errors of the estimated values. The average angular error is about 2°. For proximal femur, the error in estimating ϕ_y is smaller than those of ϕ_x and ϕ_z because rotation of the proximal femur about the Y-axis (shaft) produces a larger difference in the projected

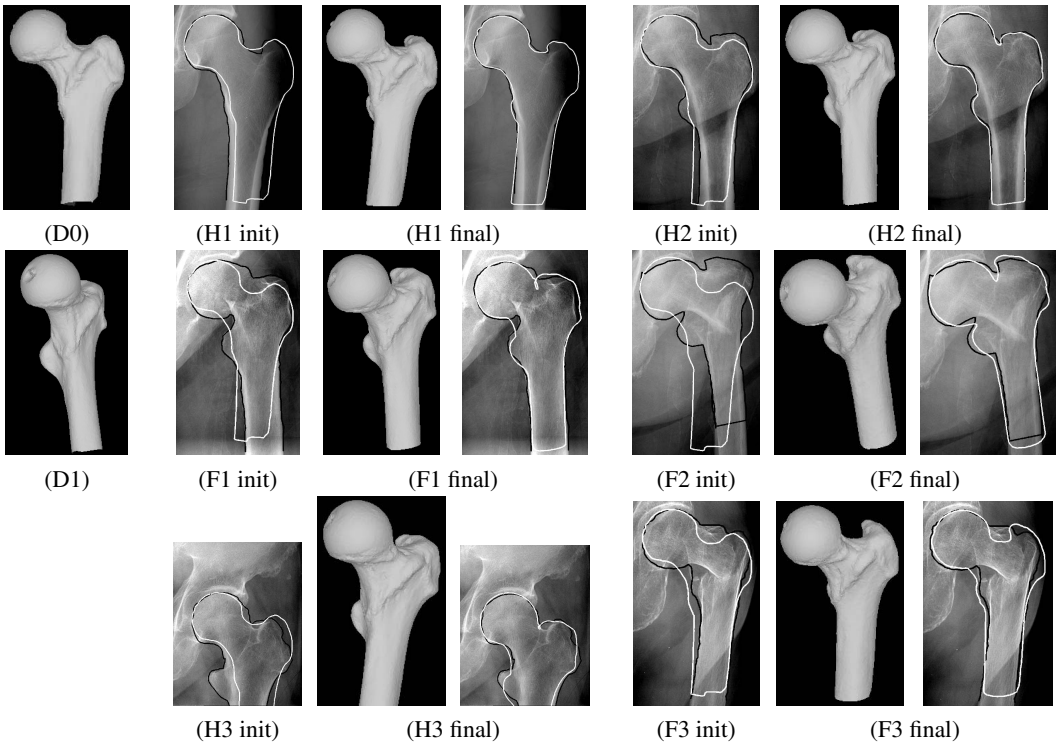


Figure 3. Sample test results. (D0) Initial default 3D pose, (D1) extreme pose (ϕ_y = 70°) for accuracy test, (H) healthy, (F) fractured. (init) Pose after automatic initialization, (final) final registered pose. Black contours: image boundary contours, white contours: projected model contours.

model contour than rotations about the X- and Z-axes. So, the algorithm is more sensitive in estimating ϕ_y. For distal femur, error in ϕ_z is the smallest for a similar reason. In summary, the algorithm can recover 3D pose accurately even when its initial default pose is very different from the actual pose.

Cross-Patient Test: A single generic 3D model is used to estimate the pose of femurs in the images of different patients. This test condition is different from those in [2, 3, 4, 5, 6, 9, 10, 11]. To generate the ground truth, approximate patient-specific 3D models were generated as follows. Healthy proximal femur images of 11 different patients were arbitrarily selected. For each image, our registration algorithm was applied to obtain a good registration of the generic 3D model with the image. The error between the projected model contour points and the image contour points was computed, and the corresponding 3D model points were displaced to minimize the error. As the 3D model points were displaced, neighboring 3D model points were also moved subject to the constraint that the surface normals at the triangular faces remained unchanged. This constraint helped to ensure that the modified 3D models were reasonably good approximations of the actual 3D models of the test images.

For each of the 11 approximate 3D models of proximal femur, 5 2D views were generated by randomly changing the orientation of the 3D model. The ranges of rotations were the same as in the single-patient test. The ranges of rotations were the same as in the single-patient test. Using the single generic 3D model, our 3D pose recovery algorithm was applied to estimate the pose of the approximate 3D model for each generated 2D view.

The algorithm successfully estimated the rotation angles of all the test cases. Table 1 tabulates the angular differences between the recovered and actual pose. Compared to single-patient test, all errors are larger because the model used for pose estimation is not the same as the actual models that generate the images. Nevertheless, the errors are still very small (about 3° or less), indicating that our algorithm can recover the 3D pose of femurs in x-ray images of different patients using a single generic 3D model.

5. Conclusion

This paper has presented an algorithm for computing the 3D pose of a bone by registering a generic 3D model of the bone to a 2D image. Test results show that the algorithm is robust against shape variations of patients. It achieves a

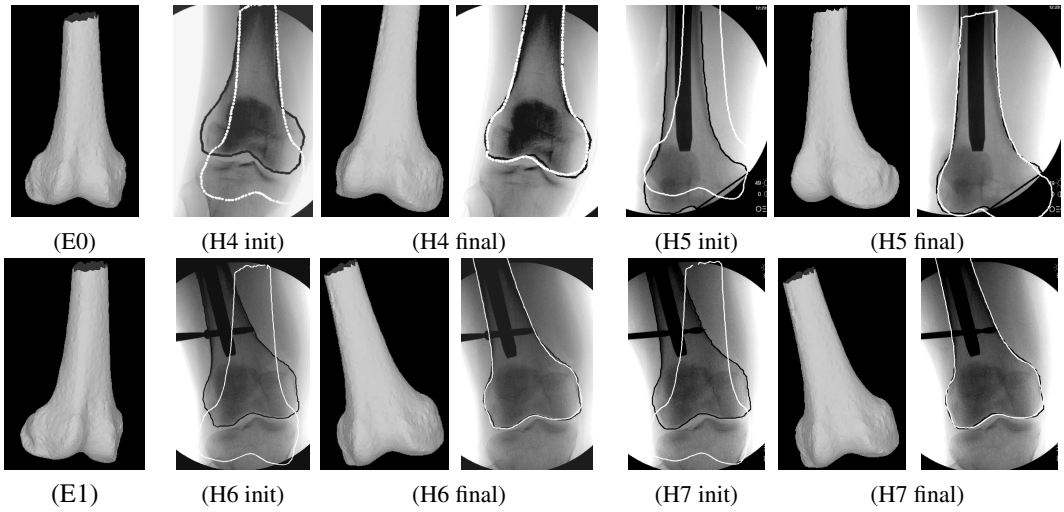


Figure 4. Sample test results. (E0) Default pose (ϕ_y = $0°$). (E1) Extreme pose (ϕ_y = $30°$) for accuracy test. (init) Pose after automatic initialization. (final) Final registered pose. Image and model contours are black and white respectively.

Table 1. Errors between recovered and actual rotation angles.

Single-Patient Test: Proximal Femur				Single-Patient Test: Distal Femur				Cross-Patient Test: Proximal Femur			
Error	ϕ_x	ϕ_y	ϕ_z	Error	ϕ_x	ϕ_y	ϕ_z	Error	ϕ_x	ϕ_y	ϕ_z
mean	2.33	1.05	1.41	mean	1.61	2.03	0.84	mean	3.21	2.48	1.58
median	2.02	1.06	1.16	median	1.08	2.06	0.66	median	2.97	2.55	0.94
std dev	1.82	0.73	1.14	std dev	1.46	1.34	0.71	std dev	2.13	0.90	1.58

93.9% success rate in registering images of both healthy and fractured femurs. Tests performed on synthetically generated images show that the algorithm can recover 3D pose of the femur from single images accurately even when the 3D model's default pose is very different from its actual pose, and partial occlusion is present in the images. The error is about $3°$ or less. Our work, thus, helps to bring 3D pose estimation method to computer-assisted diagnosis and surgery that do not routinely involve CT or MR volume images.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. on PAMI*, 14(2):239–256, 1992.

[2] A. Czopf, C. Brack, M. Roth, and A. Schweikard. 3D-2D registration of curved objects. *Periodica Polytechnica Ser. El. Eng.*, 43(1):19–41, 1999.

[3] A. Guéziec, P. Kazanzides, B. Williamson, and R. H. Taylor. Anatomy-based registration of CT-scan and intraoperative x-ray images for guiding a surgical robot. *IEEE Trans. on PAMI*, 17(5):715–728, 1998.

[4] D. A. LaRose. *Interactive X-ray/CT Registration Using Accelerated Volume Rendering*. PhD thesis, Carnegie Melon University, USA, 2001.

[5] S. Lavallée and R. Szeliski. Recovering the position and orientation of free-form object from image contours using 3D distance maps. *IEEE Trans. on PAMI*, 17(4):378–390, 1995.

[6] H. Livyatan, Z. Yaniv, and L. Joskowicz. Gradient-based 2D/3D rigid registration of fluoroscopic X-ray to CT. *IEEE Trans. on MI*, 22(11):1395–1406, 2003.

[7] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[8] M. Overmars, M. D. Berg, M. V. Kreveld, and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, Berlin, 1997.

[9] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2-D–3-D medical image registration. *IEEE Trans. on MI*, 17(4):586–595, 1998.

[10] E. B. van de Kraats, G. P. Penney, D. Tomaževič, T. van Walsum, and W. J. Niessen. Standardized evaluation methodology for 2-D–3-D registration. *IEEE Trans. on MI*, 24(9):1177–1189, 2005.

[11] L. Zöllei, E. Grimson, and W. Wells. 2D-3D rigid registration of x-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators. In *Proc. of CVPR*, volume 2, pages 696–703, 2001.